Journal of Forestry Research (2013) 24(3): 503–508
DOI 10.1007/s11676-013-0381-6

ORIGINAL PAPER

Development of amplified consensus genetic markers in *Taxodiaceae* based on *Cryptomeria japonica* ESTs data

LU Yong-quan • JIA Qing • TONG Zai-kang

Received: 2012-03-25; Accepted: 2012-08-22

Abstract: Amplified consensus genetic marker (ACGM) is a PCR-based marker technique that uses primers designed within conserved regions of coding sequences. After a comparison of *Cryptomeria japonica* and *Arabidopsis* ESTs to search for conserved sequences, 237 single e-PCR products were obtained. We randomly selected 110 candidate ACGM markers to test. Of the 110 candidate ACGM markers tested, 106 yielded stable and clear PCR products in *C. japonica*. We then tested the utility of these 106 primer pairs in 10 species, representing 7 genera of *Taxodiaceae*. The number of specific amplification primer pairs among those 10 species varied from 49 to 103 (or 46.2~97.2%). The 106 primer pairs (ACGM loci) were high transferable to *Cryptomeria fortunei* Hooibrenk (97.2%) but were low in *Metasequoia glyptostroboides* (46.2%). The number of PCR bands per primer pair ranged from 1.06 to 1.15, which means that most of the ACGM primers can obtain a single band within these 10 Taxodiaceae species. In summary, our study shows that ACGM is a technique applicable for marker development even in species with limited sequence data.

Key words: ACGM; *Taxodiaceae*; *Cryptomeria japonica*; EST database

Introduction

Molecular markers have been widely applied in the construction of genetic maps and gene mapping, molecular marker-assisted selection during breeding, gene cloning, and comparative genomics. Many types of molecular markers have been developed since restriction fragment length polymorphisms (RFLP) were developed more than three decades ago (Botstein et al. 1980). To date, the more powerful and available markers are those based on polymerase chain reaction (PCR) techniques. Briefly, there are two types of PCR-based markers. One is random primer markers which can be used in most species, examples are randomly amplified polymorphic DNAs (RAPDs) (Williams et al. 1990) and amplified fragment length polymorphisms (AFLPs) (Vos et al. 1995). The other is locus-specific primers markers such as simple sequence repeat (SSR), which must be developed from and used in target species (Becker et al. 1995). The flexibility of random primer markers is unquestioned since they can be used in nearly all species; however, their reliability, especially of RAPDs, is in doubt to some extent. Although the utility of SSRs in genetics studies is well established, the isolation and characterization of such markers via traditional methods are costly and time consuming, making the *de novo* development of SSRs unrealistic for some taxa (Pashley et al. 2006).

Amplified consensus genetic marker (ACGM) is a PCR-based marker technique which focuses on primer design within conservative regions of coding sequences (Fourmann et al. 2002). By comparing regions of homology between closely-related species, it is possible to develop ACGM markers for those species that have sufficient sequence data (Lu et al. 2006a). Expressed sequence tags (ESTs) are generated from single-pass sequencing of randomly picked cDNA clones (Adams et al. 1991). For those species with no whole genome sequence availability, ESTs offer the best solution for consensus region analysis of expressed genes within closely-related species. Thus, the use of EST databases to develop ACGMs for target species is an inexpensive and rapid alternative to using traditional methods for the development of locus specific molecular markers (Wang et al. 2005).

At present, however, there remain many species with no EST data. Yet, it is often research in these species (with little available sequence data) that would benefit most from more efficient molecular marker development to support research into their biology. According to Yang et al. (2007), EST-derived markers are likely

Fund project: This work was funded by the Natural Science Foundation of China (30800879) and project 2009R50035 supported by Forest Seedling Industry Innovative Team of Zhejiang province in China.

The online version is available at http://www.springerlink.com

LU Yong-quan (✉) • JIA Qing • TONG Zai-kang
The Nurturing Station for the State Key Laboratory of Subtropical Silviculture, Zhejiang Agriculture and Forestry University, Lin'an 311300, China. Tel: 86+571+63743855; Email: luyongquan@126.com

Corresponding editor: Hu Yanbo

to be conserved across a broader taxonomic range than any other type of marker. Therefore, for those species with a scarcity of sequence information, EST data from closely related species could be used. For example, Lu et al. (2006b) developed ACGM markers in *Gramineae* using rice databases from the cultivars Nipponbare and 93-11. These markers were transferable among several grass species.

The *Taxodiaceae* family within the gymnosperm class are relict plants from the Cretaceous period (Wu et al. 1998). The family was an important component in forest vegetation of the northern hemisphere from the late Cretaceous to the mid-tertiary era approximately 115 to 30 million years ago. In the late tertiary and pleistocene, however, the family underwent a widespread reduction resulting in the present day relictual genera with restricted distributions (Sehlarbaum and Tsuehiya 1984).

Presently, research within the *Taxodiaceae* lags behind even other tree species. Further, there is limited sequence data available from *Taxodiaceae* species. The lack of sequence data makes it difficult to develop locus specific primers for these species and the choice of markers for genetic and genomic research in these species is narrow. Thus, it will be useful and necessary to develop and test additional markers for *Taxodiaceae* species to support further research. *Cryptomeria japonica* has the most available sequence data of all species of *Taxodiaceae*. By comparing *C. japonica* ESTs with *Arabidopsis* whole genome coding sequences to search for conserved regions of homology, we exploited the ACGM technique for marker development in the *Taxodiaceae*. This paper provides a case study of the utility of available *C. japonica* EST resources for the development markers necessary for the genetic analysis of members of the *Taxodiaceae* family.

Materials and methods

Search for putative ACGMs

In total, 56,646 EST sequences of *C. japonica*, released by the Plant Genomics Database (GDB, http://www.plantgdb.org) were downloaded. In addition, the coding DNA sequence (CDS) data of *Arabidopsis* (*Arabidopsis thaliana* ecotype Columbia) were downloaded from The Arabidopsis Information Resource (http://www.arabidopsis.org/).

We developed a pipeline using Perl script to search for conserved regions of homology between *C. japonica* and *Arabidopsis*. The initial step was to identify the conserved regions within ESTs of the available sequences in *C. japonica* by aligning the EST sequences of *C. japonica* with the CDS of *Arabidopsis* using BLASTN.30. A *C. japonica* EST was thought to be homologous to an Arabidopsis CDS only if there were at least 200 bp overlapping and 80% similarity between them. We then identified the possible varied locus within the ESTs of *C. japonica* by aligning the EST sequences with the CDS of *Arabidopsis* using BLAT. In order to include additional putative polymorphisms, putative ACGM should be those EST sequences containing at least one varied region. The third step was to design primers for *C. japonica* ESTs containing possible positions of conserved region. For each of the EST, a pair of primers was designed using program ePrimer3 (Rozen and Skaletsky 2000) on a 200-bp sequence cut from the *C. japonica* EST with 100 bp on each side of the target varied locus.

Targeting of candidate ACGMs by electronic PCR

The designed primers were tested by electronic PCR (e-PCR) on the EST sequences of *C. japonica*. To increase the quality and usability of the *in silico* predicted ACGM markers, we required exact matches between primers and templates and set a 300-bp margin for the product size for the e-PCR. We accepted a putative ACGM locus as a candidate ACGM marker only if those primers successfully and uniquely amplified the correct target in the e-PCR. The candidate primers were selected and named with the abbreviation TA (for *Taxodiaceae* ACGM) followed by a unique number (e.g. TA 11).

Experimental verification and evaluation of ACGM markers in *C. japonica* by PCR

Fresh leaves of *C. japonica* were collected from Hangzhou Plant Garden (Zhejiang, China). Total genomic DNA was isolated from 200 mg of fresh ground leaf tissue using the CTAB method according to Murray et al. (1980).

All primers used were synthesized by the Nanjing Jinsite Biological Engineering & Technology Company (Nanjing, China). PCR was performed in 20 μL reactions containing 50 ng of template DNA, 0.5 μmol/L of each primer, 200 μmol/L of each dNTP, 1.5 mmol/L of $MgCl_2$, 1 unit of Taq polymerase, and 2 μL of 10 × PCR reaction buffer. A touchdown PCR program (Don et al. 1991) was used: 5 min at 95°C; 10 cycles of: 30 s at 95°C, 30 s at 58°C minus 0.3°C per cycle, 1 min at 72°C; 20 cycles of: 30 s at 95°C, 30 s at 55°C, 1 min at 72°C; and 7 min at 72°C for a final extension. For those primer pairs that did not generate good amplification results, the initial annealing temperatures were adjusted from 55°C to 60°C. Each of the primer pairs was tested twice to confirm the repeatability of the observed bands in *C. japonica*. PCR products were separated on 1% agarose gel. Gels were stained with ethidium bromide to visualize DNA bands.

Experimental evaluation of ACGM marker development in other *Taxodiaceae* species

To evaluate the transferability of ACGMs in *Taxodiaceae*, genomic DNA from 10 species (Table 1), representing seven genera of *Taxodiaceae*, were collected from Hangzhou Plant Garden (Zhejiang, Chian). Genomic DNA were extracted using the CTAB method as described in 2.3. PCR reactions were the same as mentioned above. Only clear and reproducible bands were counted as specific amplification. Based on the PCR data, we evaluated the allelic diversity of each ACGM marker using the polymorphism information content (PIC) value defined as $PIC_i = 1-\sum P_i^2$, where P_i is the frequency of the i*th* marker.

Table 1. Accessions used for ACGM analysis

Accession number	Species	genus	Chromosome number*
1	*Taxodium distichum*	*Taxodium*	2n=22
2	*Taxodium ascendens* Brongn	*Taxodium*	2n=22
3	*Sequoia sempervirens*	*Sequoia*	2n=66
4	Variety of *Sequoia sempervirens*	*Sequoia*	unknown
5	*Cryptomeria fortunei*	*Cryptomeria*	2n=22
6	Variety of *Cryptomeria fortunei* Hooibrenk	*Cryptomeria*	unknown
7	*Cryptomeria japonica*	*Cryptomeria*	2n=22
8	*Metasequoia glyptostroboides*	*Metasequoia*	2n=22
9	*Cunninghamia Lanceolata*	*Cunninghamia*	2n=22
10	*Taiwania cryptomerioides*	*Taiwania*	2n=22
11	*Glyptostrobus pensilis*	*Glyptostrobus*	2n=22

All plant materials were collected from Zhejiang plant garden in China. *data from Schlarbaum and Tsuchiya (1984)

Results

Candidate ACGM markers

After screening 56,646 EST sequences from *C. japonica* with the developed Perl pipeline, we obtained 10,289 ESTs carrying putative ACGM loci. We successfully obtained 1,705 (17%) e-PCR products from the putative ACGM loci. That some sequences did not yield a product might be due to the stringent conditions for e-PCR. Within the successful e-PCR products, there were multiple ACGM loci that obtained the same PCR product. Multiple-gene-copy markers are not desirable for genetic studies, so we discarded these putative ACGMs and their primer pairs. After filtering out these redundant ACGM loci, we successfully obtained 237 single e-PCR products and 110 of these were randomly selected as candidate ACGM markers.

Experimental tests of candidate ACGM markers

Following the *in silico* analysis, the 110 randomly selected candidate ACGM markers were tested in *C. japonica* experimentally by PCR (Fig. 1). Of the 110 candidate ACGM markers tested, 106 yielded a single, clear PCR product in *C. japonica*. Theoretically, since all the primer pairs were designed using *C. japonica* ESTs, all primer pairs should amplify products in the *C. japonica* genome. We suspected that the reasons for amplification failure of the four remaining candidate loci might have been variation of the primer location within the *C. japonica* used in this study, or sequencing errors in the EST data.

Transferability of ACGM primers within the *Taxodiaceae* genera

The utility of the 106 PCR-confirmed ACGM markers in various *Taxodiaceae* species was tested (Fig. 2). The number of specific amplification primer pairs among the 10 species varied from 49 to 103 (or 46.2%–97.2%, Table 2). The 106 primer pairs (ACGM loci) were highly transferable to *Cryptomeria fortunei* Hooibrenk (97.2%) but less transferable to *Metasequoia glyptostroboides* (46.2%). The PCR bands per primer pair varied from 1.06 to 1.15, which means that most of the ACGM primers obtained single band within the 10 *Taxodiaceae* species. The PIC values of the markers varied from 0 to 0.50 with an average of 0.27. The polymorphism level of the ACGM marker was generally not high. However, a higher estimate of the polymorphism level of ACGM markers could probably be obtained if methods of DNA fragment analysis with higher resolution capacity (e.g. denaturing PAGE, usually used for SSR analysis or DNA sequencing) were adopted. The primer sequences and putative function for the 106 new ACGM markers are shown in Table 3.

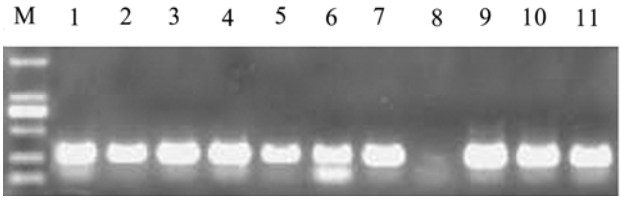

Fig. 1 Primers screened in *C. japonica* separated by electrophoresis on 1% agrose gel (M representing Marker DL2000). Number 1 to 11 representing sample of variety of *C. fortunei Hooibrenk*, variety of *S. sempervirens*, *G.pensilis*, *T. cryptomerioides Hayata*, *C. japonica*, *C. fortuneiHooibrenk*, *C. Lanceolata*, *M. glyptostroboides*, *T. ascendens. Brongn*, *S. sempervirens*, *T. distichum*, respectively.

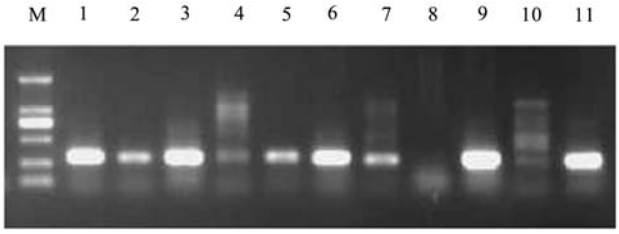

Fig. 2 Primers screened in *Taxodiaceae* separated by electrophoresis on 1% agrose gel (M representing DNA Marker DL2000). Explanations of number 1 to 11 are the same as Fig. 1.

Table 2. PCR results of 106 pairs of ACGM primers in species of Taxodiaceae

Material No.	Specific amplification primer pairs		Bands of PCR product	
	Number	Percentage	Number	Bands per primer pair
1	73	68.9	80	1.1
2	83	78.3	91	1.1
3	53	50	56	1.06
4	62	58.5	67	1.08
5	103	97.2	115	1.12
6	98	92.5	111	1.13
8	49	46.2	56	1.14
9	67	63.2	77	1.15
10	74	69.8	84	1.14
11	84	79.2	97	1.15
Mean	68	64	---	1.02

Table 3. Sequences and related information of 106 ACGM primers. The function was presumed from conserved regions of homology between *C. japonica* and *Arabidopsis*.

Marker	Forward primer(5'-3')	Reverse primer (5'-3')	Gene function or putative function
TA1	ACGCCAAGAGAGTCACCATC	GCTTAAACAAAAAGCAATCATTAGG	Histone superfamily protein
TA2	CGAACCTGTTGTTTTTGATACA	CACCAGTCACAGCAATCAGC	Chlorophyll A-B binding family protein
TA3	CCAAATGGTGGGTGTTGAA	CCTGCCACGTTTTGATTCTT	Pectin lyase-like superfamily protein
TA4	GAAATGCGATGAAAGGGAAA	TGTGATGCAGGTCCATTATGA	MYB97, AtMYB97
TA5	GAGACCTCTGAGGCAGATCG	TGGAATTGGATGCTTCCTTC	Glycolipid transfer protein (GLTP) family
TA6	TGTTGGACATCTGGACGAAA	TCCATGACACAAAAGTCACCA	Ribosomal protein S21e
TA7	GGTGTAGTCAACTGGCGTTTG	TGATTTTAGTAGAAAGATGTGGGTGT	transposable element gene
TA8	CAACACCAATGGGAGCAACT	CTGCCCTTTTAACTGTCATGC	unknown protein
TA9	TGACCAAGGTTGACAGACGA	CACCACTTGGCTCCTTCTTC	GTP binding Elongation factor Tu family protein
TA10	CGAGTCTCTATCCCGTCGAG	GGGTGTTTGACACGAACCTT	Glyptostrobus pensilis internal transcribed spacer 1
TA11	CGGAAGAAGAGGCGTAAGAA	AATGGCAAAAAGCCACAATC	transposable element gene
TA12	TGTTGGGGCTTTGCTAGTTT	TCCTGTGAATCTCACCCAGA	ATRAB11C, ATRABA2A, ATRAB-A2A, RAB-A2A, RAB11c
TA13	GCCTGTAGAAGCAGGCTTGA	AATATCCACAGCCACATCTGC	RAB18, ATDI8 | Dehydrin family protein
TA14	CGGGGGCGGTTTATATACTT	TCATGCATAGAATTTGGGAGAA	cytochrome c biogenesis protein family
TA15	GGCGTTCAAGCTAGAAGATGA	TTTTTCGTAATTTCTTTTACGATGAT	transposable element gene
TA16	TGGTAAAAGCAAAAGACGAAA	GCCCTTAATGGCCTGATGTA	SYTD, ATSYTD, NTMC2TYPE2.2, NTMC2T2.2, SYT4
TA17	ATCGGCGCCTTATGGTAACT	GGCTTCCGCATTATTTTCAA	Homeodomain-like superfamily protein
TA18	CCTTCGGCTTGAAAAACAAG	CTTTTCACCCTTCCGTTGAA	Zinc finger C-x8-C-x5-C-x3-H type family protein
TA19	AACCCTCACTGGAAAGACCA	ATCTTGGCCTTGACGTTGTC	polyubiquitin 10
TA20	GGGGAAAGAAGACCCTGTTG	CACGAGATTTCTGTTCTCGTTG	Glyptostrobus pensilis 26S ribosomal RNA gene
TA21	TTCACATTGGAGTCGGTCAA	AGGGCTCTGCTGTTCAATGT	S-adenosylmethionine synthetase 1
TA22	GCAATCTAAGCGTCATGCTG	TTTTTCACCTTTCCCTCACG	chloroplast-encoded 23S ribosomal RNA
TA23	TTACGGTTTCGAGGGTCATC	CCTCATTGTCCCAACCAGAC	NADH dehydrogenase subunit 9
TA24	GATTACGGTTTCGAGGGTCA	TTTGAAGGAGCCAGTTGGAG	NAD9
TA25	GGGCGACTGTTTACCAAAAA	CTTTCTGTCCAGGTGCAGGT	chloroplast-encoded 23S ribosomal RNA
TA26	GGGAGAGCAATCTAAGCGTCA	TTTTTAGCTATCGGTCACTCAGG	chloroplast-encoded 23S ribosomal RNA
TA27	GCAATCTAAGCGTCATGCTG	TGCTCACGGTACTATTTCGCTAT	chloroplast-encoded 23S ribosomal RNA
TA28	ATGGTCACATCTTCGGCTCT	CGGACCACCTCGTAATCATC	Protein kinase superfamily protein
TA29	TCCATCGCAATCCTTTTCAT	CGGTACCTGAGCCATTTCAT	2-oxoglutarate and FeII-dependent oxygenase superfamily protein
TA30	CTCAGATGTGTGGGCAAAGA	CGGATGAGTTCCGAATGTTT	CCCH-type zinc finger protein with ARM repeat domain
TA31	CCTAGAGGCACTGGGTTCAA	GGCAGTGAGGGTCTCGATAA	S-adenosyl-L-homocysteine hydrolase
TA32	ATTTTCAGGGCCAGATTCCT	CCATTCTGCCGTTACAATGAA	transposable element gene
TA33	ATGCAGTATGGGGATTGCAG	GAGCCTCGACCCCTTAAGAT	NAC domain containing protein 47
TA34	GCAGCGTGCAGTGTGTATG	GAAAAATCAATCCAAGCCACT	alpha-1 tubulin
TA35	GGAGGAGCTAGACGCAGAGA	GTCCAGAAGGCCGTATGTGT	Ribosomal protein S4
TA36	AAACTGCGGATGGCTCATTA	CGAACCCTAATTCTCCGTCA	rRNA
TA37	GTATGGCCGGCTAATTCGAG	AGCCAATTAAGGCCAGGAAC	Symbols: rRNA
TA38	TTATCAGGTCGCAGCGGTA	AGCCAATTAAGGCCAGGAAC	Symbols: rRNA
TA39	AATCCCTTAACGAGGATCCATT	AGCCAATTAAGGCCAGGAAC	Symbols: rRNA
TA40	GAGCCCAAAATCAGACAAGC	TCACCCTCCTTTGCTTCATC	HEAT SHOCK PROTEIN 89.1
TA41	GGACGCCAAGAGAAGAAAGA	ACACACCACGCATACAATCC	cold-regulated 47
TA42	ACTTCACGGAGCACAAAGG	GCAATTGTAACAGCGCCTAA	Histone superfamily protein
TA43	TGGGTACATTCTCGAGGGTAA	GAGCATCCATTCATCTCTCAAA	Ribosomal protein S8e family protein
TA44	AAGGCACATTCTGAGCGAGT	TGTGGCAATTTATTGCATCATT	Chalcone and stilbene synthase family protein
TA45	GGTTGCATCAGCTTGAGATGT	TGCGAAGATGCATGTTGTTT	transposable element gene
TA46	GGCTGGAATTGTGAAGAAGG	GGGAAACAGCGCAATGAT	Translation initiation factor SUI1 family protein
TA47	AAGATTGTCAGTGGCGTGTG	GTCTACGCCAGCTCTCCTTG	Ribosomal protein L32e
TA48	GACAAGGGAGGTTTGATGGA	AAACAACATCCACAGCCACA	unknown protein
TA49	GGCCTAGGTGTAACATATTGGA	AATGAGCAGCCATGGGAGT	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
TA50	TTCAACGTGCTTGCTTGTTC	AATCTTCAAGCTTGGCCGTA	proteasome alpha subunit D2
TA51	TGAAAACAGCTCAATTTGCAC	CACTAAGCTTGTCGCCACTG	pseudogene, similar to putative carboxyl-terminal, proteinase,
TA52	TGAACCGAAAGCCGAGTTAC	TGCTACTACCACCAAGATCTGC	Glyptostrobus pensilis 26S ribosomal RNA gene

Continued to Table 3

Marker	Forward primer(5'-3')	Reverse primer (5'-3')	Gene function or putative function
TA53	CGTCACGTAAGTGGGTTGTG	AGTTGTTCTGAACCGGCATT	Pinus strobus homeobox transcription factor KN2 (Kn2) gene
TA54	GGATGCTTACCGAAAGAAGG	TCAATGAATTGGTTGCATATGAT	Ribosomal protein S7p/S5e family protein
TA55	GCAATCTAAGCGTCATGCTG	TTGCCCACGGTACTATTTCG	chloroplast-encoded 23S ribosomal RNA
TA56	GTTGTGAAGGGGCAAAAATG	GGCTTCCTGGCAGTACAAAC	cell elongation protein / DWARF1 / DIMINUTO (DIM)
TA57	TAGAGTGCAAGAAGGCGTTG	ATTCCTTTTCACCCCCAATC	Pectin lyase-like superfamily protein
TA58	AGGGAACGGGCCCTTATATC	TTTTCGTTCCTAGTCGTCATGG	unknown protein
TA59	TGTTGCAGGGTTTGTGTAGG	TGGATTCCCACAGTTGAACA	plasma membrane intrinsic protein 1;4
TA60	ATCTGGGTGGAGGGACTTTT	GAGTCGATCTCGATGGTGGT	heat shock cognate protein 70
TA61	AGGACGCTTGTTTCTCTGAATC	AGGGTGGACTCCTTTTGGAT	ubiquitin 13
TA62	GGGTTTTGTTATACGAGCCTGT	TGGACTCCTTCTGGATGTTG	Symbols: UBQ13
TA63	CAAACCAAAACCAGGCTCTC	CGGAAGAGGAGAATCAGTGG	Oxidoreductase, zinc-binding dehydrogenase family protein
TA64	CGAAGGATTTGGGGTTTGTA	GGTCATAAAGCGGACGAGAA	KNOTTED1-like homeobox gene 4
TA65	TTGAGGCTGGGTTTTACAGG	CAGGCTGGTGCATCCTAATA	unknown protein
TA66	CTCACTTGTCTGTATTGGAACGA	CGGCGTATCATCCAACAGTA	Peroxidase superfamily protein
TA67	TGTCATAGGAGGTGAAAATTTGAT	CAACCGTTTTACGATAACAGCA	Translation initiation factor SUI1 family protein
TA68	AGGAGTCCACCCTTCACTTG	CGTCCATCCTCAAGCTGTTT	Ubiquitin family protein
TA69	ATGGCAGAAAAGGCAATACG	CGCCTCAACTTCGACTGAA	Symbols: ATEXPB1, EXPB1, ATHEXP BETA 1.5
TA70	CTGGTGAGGCACGTTCTGT	TCTCCTCATTCAAAGGATCG	O-Glycosyl hydrolases family 17 protein
TA71	GAGCTTGGCAAGGTTATTCAA	CTGCTGCTCACACTGTCCAT	Translation initiation factor SUI1 family protein
TA72	GCGTACCATTTTGATTGGAGA	CAAAACAGATTGGCTGCCTTA	unknown protein
TA73	CTTTAGCACTCAGCGATCATTT	CAGCGGTTGACTGGATCTTC	chloroplast-encoded 23S ribosomal RNA
TA74	GAGCTCTGACACCATCGACA	CCGTTACGGCTCAAACATAAA	polyubiquitin 10
TA75	GAGAAGAAGGCAATGGTGGA	CAAGGCCTGTCCCTGAAATA	Remorin family protein
TA76	AGCAGTTCGGATAGTGACACC	TGCTGTGGAATCCAAATTGA	Leucine-rich repeat transmembrane protein kinase
TA77	GCAAGACCATTACCCTGGAA	GGCCTACGAACATCGAACAT	ubiquitin 4
TA78	ACGAACTCTTCGAGGCAAAA	TGCAGAGTCTCGGTTCACAG	unknown protein
TA79	GCTTAGTGTGCGACTCGTTG	TGCTCTTTGGGGGTGATTTA	pseudogene, hypothetical protein
TA80	TTGCTGCCGTTTTGGTAAAT	CCCAATTCTCTTCAACTTCACA	Transducin/WD40 repeat-like superfamily protein
TA81	GGATAGAGGAGGAGGAAGACG	CACTGTAATGTCCCCGATGG	unknown protein
TA82	CCGCAGCTCAATCATTGTAA	CCTCTGCAGAACAAAAGTTGC	CAM8, AtCML8
TA83	AGGCATACAGGCAGCAGAAT	GGTATCGTGTGTGGATCATTTG	WUS, PGA6, WUS1,Homeodomain-like superfamily protein
TA84	TAGCGGGACGTTCTGTGAAT	GGCAACTTCGTCACTTTCGT	mitochondrial 26S ribosomal RNA protein
TA85	TGATTAAACAAGCGCAAATCA	GTGAGGCGGCATGAGAATA	unknown protein
TA86	CGAAAGGGAATCGGGTTAAT	GACCAGAGGCTGTTCACCTT	unknown protein
TA87	AATCCCTTAACGAGGATCCATT	AGCCAATTAAGGCCAGGAAC	rRNA
TA88	CTTCTGCCCCCAAGACCTA	TCCGGTTTACCTGTCCAATC	vacuolar protein sorting 41 |
TA89	GCATTGAAATTTTCTGCTTTCTG	ATGAGCCTCTTGGCATCAAA	Heat shock protein 70 (Hsp 70) family protein
TA90	TTGGTGCACATTTTCGTGTT	AGAAGTTGGTGGCTGTGACC	ATEXPA8, EXP8, ATEXP8, ATHEXP ALPHA 1.11, EXPA8
TA91	ATTGCAGTTGCAGCCTTTCT	ATTCCCACGGTTGAACACAT	plasma membrane intrinsic protein 1B
TA92	CTGTAAGCAAAGCCCCTCTG	ACCAGAAGGAGGAGGAGGAG	pseudogene, glycine-rich protein
TA93	GCTGCAAACGAATGAATGTC	TATTTGCCCTGCGAAGAGAG	unknown protein
TA94	CAGCACCAAGAGACAACCTG	TGGCAGAGCAAATCATGTTC	Translation protein SH3-like family protein
TA95	TAGAGAAAGGGACCGTGACC	CATAGGCCAAAAACTAGTGTGC	unknown protein
TA96	AGGAAAGGCCGAGGTTTTAC	GAGTCTCGGCTGCAACCTTA	pseudogene, similar to fiber protein Fb12,
TA97	TCAGCGAAAGAGAAGAAGGAA	CATGGCGACCTGGTTCTTAG	Heat shock protein 70 (Hsp 70) family protein
TA98	GCGGAGGCTTTGATTTACAG	AGTTTTGCCCGTTAGGGTTT	UBQ10
TA99	AATTCTTCGCGGTCTCAGAA	TGTGCAATTTCTCGAACCAA	Histone superfamily protein
TA100	GTGTGTTGGGTTGTTTGTGC	TCCAGAGAGCAATATCGATGG	GTP binding Elongation factor Tu family protein
TA101	CAGCGATTTTGAAGTGTGTGA	TTTGCATGAATGAAGTCTTTGG	transposable element gene
TA102	TGAAGGGGAATATGACGATGA	CAATGTAACGTAGCGGAGAGA	tubulin beta-7 chain
TA103	ATCCTCCTCTGGGACGTTTT	CACTATGATGCCACCGTCTG	GTP binding Elongation factor Tu family protein
TA104	CACTGACAAGCTTGGTTTGC	AGCCAGACCCACTACAAAATG	ADP-ribosylation factor A1E
TA105	TGTCAAGATTGATTGCGTCTG	GTTTGTTGTTGCGGTTGTTG	RING-H2 group F1A
TA106	GTGTGGAAGGTTGGAATGGT	TGCTCGGTGCACAATCATTA	phenylalanine ammonia-lyase 2

Discussion

Molecular markers are useful tools for genetics and breeding research. In recent years, high-density oligonucleotide microarrarys and next-generation sequencing technologies have resulted in a considerable increase in the amount of available genome sequence data (Brady et al. 2009). However, for many species, especially woody plants, publicly available sequence information remains limited. At present, it is not yet feasible to sequence the genome of a particular species under study without multi-laboratory or multi-national efforts. However, as we verified in this paper, EST data can work together with available whole genomic sequences to develop genetic tools for taxonomic groups closely related to the EST sets. The primers for our putative ACGM loci were verified in several species within the *Taxodiaceae*. According to these results, it is now possible to follow similar strategies to develop primers in a variety of species for which sequence data are few, making ACGM marker development a reasonable choice for these species.

Following the ACGM procedure, primers are designed within conserved homologous regions of coding sequences. Therefore to exploit ACGM to a greater extent the reference sequence, in our case *Arabidopsis*, should be as closely related as possible to the chosen experimental species that the EST data represents. Logically, it would have been better to use the sequence data from a woody plant such as poplar than *Arabidopsis*. The draft genome sequence of poplar has been completed and is available (GDB). However, when we attempted to blast the EST sequences of *C. japonica* against poplar data, two issues arose. First, the outcome was no more useful than that for *Arabidopsis* alone (i.e. no major increase in homologous regions was seen, data not shown). Second, the annotation for poplar coding sequences is not as developed as it is in *Arabidopsis*. This annotation problem would make further aspects of the study more ambiguous. In view of these issues we recommend to use *Arabidopsis* as a reference sequence when searching for consensus regions when undertaking ACGM marker development in selected woody species such as those of the *Taxodiaceae*.

Using ACGM for marker development is, however, not without its disadvantages. ACGM are genetic markers residing in gene sequences and they can directly reflect aspects of variation within those genes. Therefore, the maps constructed with ACGM markers could be especially valuable for genetic studies but levels of polymorphism could be low because these ACGMs are in expressed regions which have more evolutionary conservation compared with primers derived from untranscribed sequences. Nevertheless, ACGM primers can be useful for comparative genomic studies precisely because they are designed in expressed regions. This technique is applicable to a wide range of species and shows a linear relationship among different genomes within a genus (Lu et al. 2006b).

Acknowledgments

The authors thank Dr Allan Feurtado from Plant Biotechnology Institute, National Research Council of Canada for helpful suggestions on the manuscript.

References

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC. 1991. Complementary DNA sequencing: expressed sequence tag and human genome project. *Science*, **252**: 1651–1656.

Anderson JA, Churchill GA, Autrique JE. 1993. Optimizing parental selection for genetic linkage maps. Genome, **36**: 181–186.

Becker J, Heun M. 1995. Barley microsatellites: allele variation and mapping. *Plant Mol Biol*, **27**: 835–845.

Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet*, **32**: 314–331.

Brady SM, Provart NJ. 2009. Web-queryable large-scale data sets for hyposis generation in plant biology. American society of plant Biologist, Avaliable at: http://www.aspb.org

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res*, **19**: 4008.

Fourmann M, Barret P, Froger N, Baron C, Charlot F,Delourme R, Brunel D. 2002. From Arabidopsis thaliana to Brassica napus: development of amplified consensus genetic markers (ACGM) for construction of a gene map. *Theor Appl Genet*, **105**: 1196–1206.

Lu YQ, Wang XS, Huang WS, Xiao TX, Zheng Y, Wu WR. 2006a. Development of amplified consensus genetic markers in Gramineae based on rice intron length polymorphisms. *Scientia Agricultura Sinica*, **39**: 433–439 (in Chinese with an English abstract)

Lu YQ, Ye ZH, Wu WR. 2006b. Analysis of the phylogenetic relationships among several species of Gramineae using ACGM markers. *Acta Genetica Sinica*, **33**: 1127–1131.

Murray MG, Thompson WF. 1980. Rapid isolation of highmolecular-weight plant DNA. *Nucleic Acids Res*, **8**: 4321–4325.

Pashley CH, Ellis JR, Mccauley DE and Burke JM. 2006. EST Database as a source for molecular markers: lessons from Helianthus. *Journal of heredity*, **97**: 381–388.

Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*, **132**: 365–386.

Schuler GD. 1997. Sequence mapping by electronic PCR. *Genome Res*, **7**: 541–550.

Sehlarbaum SE, Tsuehiya T. 1984. Cytotaxonomy and Phylogeny in Certain Species of *Taxodiaceae*. *Syst Evol*, **147**: 29–54.

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res*, **23**: 4407–4414.

Wang XS, Zhao XG, Zhu J, and Wu WR. 2005. Genome-wide Investigation of Intron Length Polymorphisms and Their Potential as Molecular Markers in Rice (Oryza sativa L.). *DNA Research*, **12**: 417–427.

Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. 1990. DNA polymorphism amplified by arbitrary primers are useful as genetic markers. *Nucleic Acid Res*, **18**: 6531–6535.

Wu ZY. 1998. Flora of China. China Science Press, Volume 7: *Taxodiaceae.*

Yang L, Jin GL, Zhao XQ, Zheng Y, Xu ZH, Wu WR. 2007. PIP: a database of potential intron polymorphism markers. *Bioinformatics*, **23**: 2174–2177.